MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-

AFOSR-TR. 79-0617

LEVEL

UNIVERSITY OF MARYLAND
COMPUTER SCIENCE CENTER

COLLEGE PARK, MARYLAND
20742

79 05 18 044

TR-745
AFOSR-77-3271A

March 1979

STATISTICAL INFERENCE THEORY
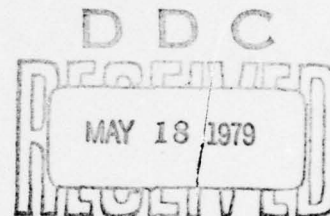APPLIED TO IMAGE MODELING

Ramalingam Chellappa
Narendra Ahuja

Computer Science Center
University of Maryland
College Park, MD 20742

D D C

MAY 18 1979

A

ABSTRACT

This paper considers the application of system
identification techniques using spectral representation
for fitting models to textures and images and consists
of two parts. In part I, we develop consistent deci-
sion rules for choosing the neighborhood in a one-
dimensional autoregressive (AR) model. In part II,
the theory is extended to the case of stationary two-
dimensional random fields.

Part I

Optimal Choice of Neighbors
for One-Dimensional Autoregressive Models

1. Introduction

We are interested in fitting an adequate model to a one-dimensional observation sequence $Y_N = \{y(1),\ldots,y(N)\}$ obtained from a stationary stochastic process. For instance, Y_N could arise from sampling image gray levels at equal intervals along a line. Y_N will be regarded as the output of a one-dimensional autoregressive (AR) process. Many decision rules are available in the literature [1,2,3] for fitting a unidirectional autoregressive (UAR) model where the current observation depends only on past ones. The problem of fitting a bilateral autoregressive (BAR) model [4], where the current observation depends on the neighbors on either side, has not been given much attention. Such models appear to be appropriate when the process is obtained by sampling an image, since for images there is no essential difference between the neighbors on one side and those on the other. In this part we consider the problem of finding the optimal neighborhood size in a one-dimensional AR model for a given empirical series.

Two approaches to modeling are the maximum likelihood approach and the Bayesian approach. We take a Bayesian approach in this paper due to the following reasons: (i) We obtain consistent decision rules for choosing the best model, and (ii) an

explicit expression for the probability density of observations given a model is obtained, which will be useful for classification of images and textures.

A comprehensive theory is available for fitting UAR models to the given data. In the maximum likelihood approach one maximizes the likelihood function separately for each model. The best model is then chosen by using the decision statistic suggested by Akaike [1].

In the Bayesian approach [2] of fitting models to the data, various plausible models (UAR of order 2, order 3, etc.) are postulated as mutually exclusive hypotheses C_i, $1 \leq i \leq r$, where r is the total number of models under consideration. The model which maximizes the posterior probability density $P(C_i|Y_N)$ is chosen as the best model with minimum probability of error. The Bayesian approach involves obtaining an expression for $p(Y_N|\underset{\sim}{\theta}, C_k)$, $1 \leq k \leq r$, where $\underset{\sim}{\theta}$ is the parameter vector characterizing the model, and then integrating this over an appropriate prior probability density function $p(\underset{\sim}{\theta}|C_k)$. A Gaussian assumption is usually made about the noise driving the model in order to obtain a simple expression for $p(Y_N|\underset{\sim}{\theta}, C_k)$. A comprehensive theory for comparison of models more general than autoregressive has been developed in [2], and the case of independent observations and linear models has been considered in [5].

In this paper we suggest a Bayesian approach for BAR model fitting. As time domain analysis is quite complicated for bilateral models we resort to spectral domain analysis. Hence instead of maximizing $P(C_i|Y_N)$, we maximize $P(C_i|Z_N)$ to obtain the minimum probability of error decision rule. Since the finite Fourier transform is a nonsingular transformation with unity Jacobian, the decision rules maximizing $P(C_i|Y_N)$ and $P(C_i|Z_N)$ are equivalent. We first write an expression for $p(Z_N|\underset{\sim}{\theta},C_i)$ using the asymptotic Gaussian properties of finite Fourier transforms and integrate it w.r.t. $\underset{\sim}{\theta}$, by using an appropriate prior probability density function $p(\underset{\sim}{\theta}|C_i)$. Using the expression for $p(Z_N|C_i)$, a decision rule that chooses a correct model with minimum probability of error is designed. Any Bayesian methodology should answer criticisms against the assumption of prior densities. In this paper, we derive $p(Z_N|C_i)$ for any arbitrary prior densities by using a theorem from the asymptotic theory of integration [10].

We show that the decision statistic suggested here reduces to the results reported in the literature for UAR models [2]. We also establish the consistency of the decision rule, i.e., the probability of choosing the jth model when the ith model is true goes to zero uniformly as $N \to \infty$.

The organization of the paper is as follows: In Section 2, we derive expressions for $p(Y_N|\underset{\sim}{\theta})$ for first order UAR and BAR models to show the relative complexities of the expressions.

The problem of fitting BAR models has not been given much attention since Whittle's work [4]. Whittle has shown how to construct UAR models that have the same autocorrelation as given BAR models, so that known procedures for UAR model fitting could be applied. But it has been pointed out that it is the multilateral scheme in general that corresponds to reality even in those cases for which the formal work of estimation, etc., is more simply performed using an equivalent UAR model.

For bilateral models the expression for the likelihood of observations is a complicated function of the coefficients, since the Jacobian of the transformation from the noise variates to observations is not unity. [In Section 2, we derive expressions for $p(Y_N|\underset{\sim}{\theta})$ for one-dimensional UAR and BAR models to illustrate the complexity of the expressions.] By considering the likelihood of transforms of observations, one obtains a simple form for the likelihood function. Specifically, using the asymptotic Gaussian properties of the finite Fourier transform, $Z_N = (z(\lambda_1), z(\lambda_2), \ldots, z(\lambda_N))$, an explicit expression can be written for $p(Z_N|\underset{\sim}{\theta})$, the dependence on the parameters appearing through the spectral density function of the process, evaluated at discrete frequencies $\lambda_1, \lambda_2, \ldots, \lambda_N$ [6,7]. Using numerical optimization algorithms [8], the maximum likelihood is evaluated, and using Akaike's criterion the best model is chosen. This procedure has been recently considered for a vector random field [9].

In Section 3 we design a decision rule that chooses a BAR model with minimum probability of error. Section 4 establishes the consistency of the decision rule. The properties of the decision rule are discussed in Section 5, and the possible applications are indicated in Section 6.

2. Expressions for $p(Y_N|\underset{\sim}{\theta})$ for UAR and BAR models

In this section we derive the expressions for $p(Y_N|\underset{\sim}{\theta})$ [11], when Y_N is assumed to obey a UAR model or a BAR model, in order to compare the relative complexities of these expressions. For simplicity, we consider first-order models.

2.1 UAR model

Given $Y_N = (y(1), y(2), \ldots, y(N))$, consider a first order UAR model

$$y(t) = \phi_1 y(t-1) + \omega(t), \quad 1 \leq t \leq N \tag{2.1}$$

where $\omega(t)$, $t = 1, 2, \ldots, N$ are identically and independently distributed Gaussian noise, $N(0, \rho)$. Consider the transformation of random variables from $\omega(1), \omega(2), \ldots, \omega(N)$ to $y(1), y(2), \ldots, y(N)$:

$$
\begin{bmatrix}
1 & & & & & \\
-\phi_1 & 1 & & & & \\
0 & -\phi_1 & 1 & & & \\
0 & 0 & -\phi_1 & 1 & & \\
\ldots & & & & 1 & \\
0 & 0 & 0 & \ldots & -\phi_1 & 1
\end{bmatrix}
\begin{bmatrix}
y(1) \\
y(2) \\
\cdot \\
\cdot \\
\cdot \\
y(N)
\end{bmatrix}
=
\begin{bmatrix}
\omega(1) \\
\omega(2) \\
\cdot \\
\cdot \\
\cdot \\
\omega(N)
\end{bmatrix}
$$

$$[\Phi_1][y(1), \ldots, y(N)]^T = [\omega(1), \ldots, \omega(N)]^T \tag{2.2}$$

The transformation in (2.2) is not exact since we have not considered the initial conditions. For large values of N, the disturbance due to initial conditions is negligible. From

$$p(Y_N | \underline{\phi}_1, \rho) = (1/2\pi\rho)^{N/2} \exp[-(N/2\rho) \tag{2.7}$$

$$(\bar{\rho} + (1/N) \sum_{t=1}^{N} y^2(t-1)(\phi_1 - \bar{\phi}_1)^2$$

$p(Y_N | \phi_1, \rho)$ is a quadratic form in ϕ_1 and its further analytical manipulation is easy.

2.2 BAR model

Given $Y_N = (y(1), \ldots, y(N))$, consider the BAR model

$$y(t) = \phi_1 y(t-1) + \phi_2 y(t+1) + \omega(t), \quad 1 \le t \le N \tag{2.8}$$

where $\{\omega(\cdot)\}$ is as in Section 2.1.

Consider the transformation of random variables from $\omega(1), \omega(2), \ldots, \omega(N)$ to $y(1), y(2), \ldots, y(N)$:

$$\begin{bmatrix} 1 & -\phi_2 & 0 & & \\ -\phi_1 & 1 & -\phi_2 & 0 & \\ 0 & -\phi_1 & 1 & -\phi_2 & \\ \cdots & & & 1 & \\ \cdots & & & & 1 \end{bmatrix} \begin{bmatrix} y(1) \\ \cdot \\ \cdot \\ \cdot \\ y(N) \end{bmatrix} = \begin{bmatrix} \omega(1) \\ \cdot \\ \cdot \\ \cdot \\ \omega(N) \end{bmatrix}$$

$$[\underline{\Phi}_2][y(1), y(2), \ldots, y(N)]^T = [\omega(1), \ldots, \omega(N)] \tag{2.9}$$

From the law of transformation of random variables,

$$p(y(1), \ldots, y(N) | \phi_1, \phi_2, \rho)$$

$$= |J| p(\omega(1), \omega(2), \ldots, \omega(N))|$$

$$(\omega(1), \ldots, \omega(N))^T = [\Phi_2](y(1), \ldots, y(N))^T$$

the law of transformation of random variables,

$$p(y(1), \ldots, y(N) | \phi_1, \rho)$$

$$= |J| p(\omega(1), \ldots, \omega(N)) | \qquad (2.3)$$

$$(\omega(1), \ldots, \omega(N))^T = [\underset{\sim}{\Phi_1}] (y(1), \ldots, y(N))$$

The Jacobian of the transformation is unity. By using the Gaussian assumption regarding $\omega(1), \omega(2), \ldots, \omega(N)$, we obtain

$$p(y(1), y(2), \ldots, y(N) | \phi_1, \rho)$$

$$= (1/2\pi\rho)^{N/2} \exp(-\frac{1}{2\rho} \sum_{t=1}^{N} (y(t)-\phi_1 y(t-1))^2) \qquad (2.4)$$

Let

$$\bar{\phi}_1 = \sum_{t=1}^{N} y(t)y(t-1) / \sum_{t=1}^{N} y^2(t-1), \text{ and}$$

$$\bar{\rho} = (1/N) \sum_{t=1}^{N} (y(t)-\bar{\phi}_1 y(t-1))^2 \qquad (2.5)$$

The exponential term in (2.4) can be rewritten as

$$\sum_{t=1}^{N} (y(t)-\bar{\phi}_1 y(t-1)+\bar{\phi}_1 y(t-1)-\phi_1 y(t-1))^2$$

$$= N(\bar{\rho}+(1/N)(\phi_1-\bar{\phi}_1)^2 \sum_{t=1}^{N} y^2(t-1)) \qquad (2.6)$$

Substituting (2.6) in (2.4) we have

3. Decision rule for BAR model selection

We are given a sequence of observations $Y_N = (y(1), \ldots, y(N))$ and r mutual exclusive compound hypotheses $C_1, C_2, \ldots, C_r$. To describe C_i, consider the stochastic difference equation $E_i(\underset{\sim}{\phi}, \rho)$

$$E_i: \quad [A_i(\underset{\sim}{\phi}, D) + B_i(\underset{\sim}{\phi}, D^{-1})] y(t) = \omega(t) \tag{3.1}$$

$$A_i(\underset{\sim}{\phi}, D) = 1 + \phi_1 D + \phi_2 D^2 + \ldots \phi_{m_i} D^{m_i}$$

$$B_i(\underset{\sim}{\phi}, D) = \phi_{m_{i+1}} D^{-1} + \phi_{m_{i+2}} D^{-2} + \ldots \phi_{m_i} D^{-(n_i - m_i)} \tag{3.2}$$

where $\quad D^r y(s) = y(r+s)$

(3.1) is characterized by an $(n_i + 1)$ dimensional vector $\underset{\sim}{\theta}^T = (\underset{\sim}{\phi}^T, \rho)$, $\underset{\sim}{\phi} \in R^{n_i}$, $\underset{\sim}{\phi}^T = (\phi_1, \phi_2, \ldots, \phi_{n_i})$, $\phi_j \neq 0$, $j = 1, 2, \ldots, n_i$ and T denotes the transpose operator. In (3.1) $\omega(t)$, $t = 1, 2, \ldots, N$ are independent and identically distributed Gaussian random variables with zero mean and variance ρ. When the coefficients in the expression $B_i(\underset{\sim}{\phi}, D)$ are all identically zero the model reduces to a unilateral autoregressive model.

We make the following assumption.

A1): The zeros of $(A_i(\underset{\sim}{\phi}, D) + B_i(\underset{\sim}{\phi}, D^{-1}))$ do not lie on the unit circle, for all i, $1 \leq i \leq r$.

Let $C = \{E(\underset{\sim}{\phi}, p); \ p > 0, \ \underset{\sim}{\phi} \in R^n\}$

$$\underline{\underline{\Delta}} \ (E, m, n)$$

C is a class standing for a set of models all having the same equation E with the same m,n but differing from each other in the numerical values of the coefficients. As long as their equations are different, the two classes are different.

From the structure of the equation in (2.9) we see that the Jacobian of the transformation is not unity, but rather a complicated function of coefficients. Also note that we do not even obtain a closed form solution for the maximum likelihood estimate $(\bar{\phi}_1, \bar{\phi}_2)^T$ of $(\phi_1, \phi_2)^T$, as in the case of UAR models.

We are interested in finding a decision rule to assign Y_N to one of the hypotheses C_i, $1 \le i \le r$. It is well known that the decision rule that chooses the true model with minimum probability of error is:

Choose hypothesis $k*$ if

$$k* = \arg \max_k \{P(C_i|Y_N)\} \qquad (3.3)$$

Consider a nonsingular transformation with Jacobian unity, $\Gamma: Y_N \to Z_N$. Then the decision rule in (3.4) is equivalent to (3.3):

Choose hypothesis $k*$ if

$$k* = \arg \max_k \{P(C_i|Z_N)\} \qquad (3.4)$$

Here

$$P(C_i|Z_N) = \frac{p(Z_N|C_i)P(C_i)}{p(Z_N)}$$

where $P(C_i)$, $1 \le i \le r$ are the prior probabilities of hypotheses and

$$p(Z_N|C_i) = \int p(Z_N|\underset{\sim}{\theta}, C_i) \; p(\underset{\sim}{\theta}|C_i) d\underset{\sim}{\theta}$$

Specifically, consider the case where Γ is the finite Fourier transform. This enables use to write an expression for $p(Z_N|\underset{\sim}{\theta}, C_i)$ by using Theorem 1 [6][7]:

<u>Theorem 1</u>: Consider the finite Fourier transform of the observations defined by

$$z(\lambda_i) = N^{-1/2} \sum_{t=1}^{N} e^{-j\lambda_i t} y(t) \qquad (3.5)$$

where $j = \sqrt{-1}$, $\lambda_i = 2\pi i/N$, $i=1,2,\ldots,N$. Let the observations obey the hypothesis C_k. For large values of N, the finite Fourier

transforms $z(\lambda_1)$, $z(\lambda_2),\ldots,z(\lambda_N)$ are independent and distributed normally with zero means and variances

$$S_{y_k}(e^{j\lambda_1},\underset{\sim}{\phi},\rho),\ S_{y_k}(e^{j\lambda_2},\underset{\sim}{\phi},\rho)\ldots\ldots S_{y_k}(e^{j\lambda_N},\underset{\sim}{\phi},\rho)$$

where

$$S_{y_k}(e^{j\lambda_i},\underset{\sim}{\phi},\rho) = \rho\ H_k(e^{j\lambda_i},\underset{\sim}{\phi})\ H_k^*(e^{j\lambda_i},\underset{\sim}{\phi}) \tag{3.6}$$

and

$$H_k(e^{j\lambda_i},\underset{\sim}{\phi}) = [A_k(\underset{\sim}{\phi},e^{j\lambda_i}) \qquad\qquad 1 \le k \le r$$
$$+ B_k(\underset{\sim}{\phi},e^{j\lambda_i})]^{-1} \qquad 1 \le i \le N \tag{3.7}$$

This theorem allows us to write an expression for the probability density of the transforms of the observations given the parameters of the model and the hypothesis it obeys.   The likelihood of the transformed observations is given by

$$\ell n\ p(\ z(\lambda_1),\ z(\lambda_2),\ldots,z(\lambda_N)|\ \underset{\sim}{\phi},\rho,C_k)$$

$$=-\frac{N}{2}\ell n 2\pi - \frac{1}{2}\left\{\sum_{i=1}^{N}\ell n\ S_{y_k}(e^{j\lambda_i},\phi,\rho)\right.$$

$$\left.+\sum_{i=1}^{N} z(\lambda_i)\,z^*(\lambda_i)/S_{y_k}(e^{j\lambda_i},\underset{\sim}{\phi},\rho)\right\} \tag{3.8}$$

Substituting (3.6) in (3.8) we have
LHS of (3.8)

$$= -\frac{N}{2}\ell n\ 2\pi - \frac{1}{2}\sum_{i=1}^{N}\ell n(\rho||H_k(e^{j\lambda_i},\underset{\sim}{\phi})||^2)$$

$$- (1/2\rho)\sum_{i=1}^{N}(||z(\lambda_i)||^2/||H_k(e^{j\lambda_i},\underset{\sim}{\phi})||^2) \tag{3.9}$$

$$= -\frac{N}{2}\ln 2\pi\rho - \frac{1}{2}\sum_{i=1}^{N}\ln ||H_k(e^{j\lambda_i},\underset{\sim}{\phi})||^2$$

$$- (1/2\rho)\sum_{i=1}(||z(\lambda_i)||^2/||H_k(e^{j\lambda_i},\underset{\sim}{\phi})||^2) \tag{3.10}$$

The structure of (3.10) is not in an appropriate form for further manipulation. We give below an equivalent expression for (3.10) in Theorem 2. Prior to that we need the following assumption:

ASSUMPTION (A2): The first and second derivatives of

$$\sum_{i=1}^{N}\ln||H_k(e^{j\lambda_i},\underset{\sim}{\phi})||^2$$

and $\sum_{i=1}^{N}||z(\lambda_i)||^2/||H_k(e^{j\lambda_i},\underset{\sim}{\phi})||^2$ w.r.t. $\underset{\sim}{\phi}$

exist for all $\underset{\sim}{\phi} \in R^{n_k}$

Theorem 2: For large values of N,

$$\ln p(z(\lambda_1),z(\lambda_2),\ldots z(\lambda_N)|\underset{\sim}{\phi},\rho,C_k)$$

$$= -(N/2)\ln f(\overline{\underset{\sim}{\phi}}_k,\overline{\rho}_k) - \frac{N}{2}((\rho-\overline{\rho}_k)\overline{d}_k +$$

$$(\underset{\sim}{\phi}-\overline{\underset{\sim}{\phi}}_k)^T\overline{Q}_k(\underset{\sim}{\phi}-\overline{\underset{\sim}{\phi}}_k) + (\rho-\overline{\rho}_k)(\underset{\sim}{\phi}-\overline{\underset{\sim}{\phi}}_k)^T\overline{S}_k + O(||\underset{\sim}{\phi}-\overline{\underset{\sim}{\phi}}_k||^3) \tag{3.11}$$

where

$$f(\overline{\underset{\sim}{\phi}}_k,\overline{\rho}_k) = \ell n\overline{\rho}_k + \frac{1}{N}\sum_{i=1}^{N}\ln||H_k(e^{j\lambda_i},\overline{\underset{\sim}{\phi}}_k)||^2+1 \tag{3.12}$$

$$\overline{d}_k = \partial^2 f(\underset{\sim}{\phi},P)/\partial P^2 \Big|_{\substack{\underset{\sim}{\phi}=\overline{\underset{\sim}{\phi}}_k \\ \rho=\overline{\rho}_k}}$$

$\overline{\underset{\sim}{\phi}}_k$ and $\overline{\rho}_k$ are maximum likelihood estimates of $\underset{\sim}{\phi}$ and ρ.

$$\overline{\underset{\sim}{Q}}_k = \frac{1}{2} \left. \frac{\partial^2 f(\phi,\rho)}{\partial \phi_i \partial \phi_j} \right|_{\substack{\underset{\sim}{\phi}=\overline{\underset{\sim}{\phi}}_k \\ \rho=\overline{\rho}_k}} \qquad (n_k \times n_k \text{ matrix}) \quad (3.14)$$

$$1 \leq i \leq n_k, \quad 1 \leq j \leq n_k$$

$$\overline{\underset{\sim}{s}}_k = \partial^2 f(\underset{\sim}{\phi},\rho)/\partial\rho\partial\phi_i \qquad\qquad (n_k \times 1) \text{ vector} \quad (3.15)$$

and

$$f(\underset{\sim}{\phi},\rho) = \ell n\rho + (1/N) \sum_{i=1}^{N} \ell n ||H_k(e^{j\lambda_i},\underset{\sim}{\phi})||^2$$

$$+ (1/N\rho) \sum_{i=1}^{N} ||z(\lambda_i)||^2 / ||H_k(e^{j\lambda_i},\underset{\sim}{\phi})||^2 \ .$$

To obtain $p(z(\lambda_1),z(\lambda_2),\ldots z(\lambda_N)|C_k)$ we must integrate (3.11) over $(\underset{\sim}{\phi},\rho)$ by using appropriate prior probability densities $p(\underset{\sim}{\phi},\rho|C_k)$. We do not make any specific assumption regarding the structure of $p(\underset{\sim}{\phi},\rho|C_k)$. The density must be regular but otherwise can be arbitrary. We use a theorem from asymptotic integration [9] to integrate over $\underset{\sim}{\phi}$ and ρ. An approximate expression for the posterior density $P(C_i|Z_N)$ is given in

<u>Theorem 3</u>: For large values of N,

$$\ell n P(C_i|Z_N) \simeq -(N/2)f(\overline{\underset{\sim}{\phi}}_k,\overline{\rho}_k) + \ell n p(\overline{\underset{\sim}{\phi}}_k,\overline{\rho}_k/C_k)$$

$$+ \frac{1}{2}(n_k+1)\ell n(2\pi/N) + (N/2)\ell n2\pi - \frac{1}{2}\ell n \ \det F_{n_k}(-g(\theta;N))$$

$$+ \ell n P(C_i) - \ell n P(Z_N) \qquad\qquad (3.16)$$

where   .

$$f(\overline{\underset{\sim}{\phi}}_k,\overline{\rho}_k) = \ell n\overline{\rho}_k + \frac{1}{N} \sum_{i=1}^{N} \ell n ||H_k(e^{j\lambda_i},\overline{\underset{\sim}{\phi}}_k)||^2 + 1 \qquad\qquad (3.17)$$

For practical applications we suggest a simplified decision rule:

Decide hypothesis is k* if

$$k* = \arg \min_{k} \{h'_k(Z_N)\}$$

where

$$h'_k(Z_N) = N\ln\bar{\rho}_k + \sum_{i=1}^{N} \ln||H_k(e^{j\lambda_i}, \bar{\phi}_k)||^2 + n_k \ln N \qquad (3.20)$$

The consistency of this decision rule can be proved similarly to that of (3.17).

The form of the decision statistics is similar to that reported in the literature [2][5]. The first two terms represent the contribution from the likelihood of transforms of the observations and the second term is due to the prior probability density function. We show that the decision statistic reduces to that reported in [2] for UAR models. For these models the first simplification is that the Jacobian $|J_k|$ of the transformation from noise variates to observations is unity. Hence [13], for the kth model

$$\ln J_k = \frac{N}{2(2\pi)^n} \int \ln||H_k(e^{j\lambda}, \underset{\sim}{\phi}) H_k^*(e^{j\lambda}, \underset{\sim}{\phi})||d\lambda$$

$$= -\frac{1}{2} \sum_{i=1}^{N} \ln||H_k(e^{j\lambda_i}, \underset{\sim}{\phi})||^2 = 0 \qquad (3.21)$$

Also the coefficients in the expression $B_i(\underset{\sim}{\phi}, D^{-1})$ are identically zero. The equations for $\bar{\phi}_k$ and $\bar{\rho}_k$ reduce to

and

$$g(\underset{\sim}{\theta};N) = -[(\rho-\overline{\rho}_k)^2\overline{d}_k + (\underset{\sim}{\phi}-\underset{\sim}{\overline{\phi}}_k)^T\underset{\sim}{Q}_k(\underset{\sim}{\phi}-\underset{\sim}{\overline{\phi}}_k)$$

$$+ (\rho-\overline{\rho}_k)(\underset{\sim}{\phi}-\underset{\sim}{\phi}_k)^T\underset{\sim}{\overline{S}}_k] \tag{3.18}$$

<u>Comments</u>: (1) we have obtained an approximate expression for $\ell nP(C_i|Z_N)$. Hence the decision rule that maximizes $P(C_i|Z_N)$ obtained here does not exactly minimize the probability of error.

(2) The expression suggested in (3.16) involves a term due to the prior probability density. The prior probabilities should be chosen to reflect the degree of knowledge we possess about the parameters. Following Jeffrey, we suggest a uniform distribution for each of the components of $\underset{\sim}{\phi}$ and a uniform distribution for $\ell n\rho$ [12].

We suggest a decision rule that approximates the minimum probability of error rule:

Decide hypothesis is k* if

$$k^* = \arg\max_{k} \{h_k(Z_N)\}$$

where $h_k(Z_N) = -(N/2)f(\underset{\sim}{\overline{\phi}}_k,\overline{\rho}_k) - (n_k/2)\ell n\ N$

$$+\ell n\ p(\underset{\sim}{\phi}_k,\rho_k|C_k) - \frac{1}{2}\ell n\ \det\ F_k(-g(\underset{\sim}{\theta},N)) \tag{3.19}$$

We establish the consistency of the decision rule in the next section.

$$\underline{\phi}_k = \min_k \{ \sum_{i=1}^{N} \frac{||Z(\lambda_i)||^2}{||H_k(e^{j\lambda_i},\underline{\phi}_k)||^2} \} \qquad (3.22)$$

and $\quad \overline{\rho}_k = \frac{1}{N} \sum_{i=1}^{N} ||Z(\lambda_i)||^2 / ||H_k(e^{j\lambda_i},\underline{\phi}_k)||^2 \qquad (3.23)$

The statistic $f(\overline{\underline{\phi}}_k,\overline{\rho}_k)$ reduces to $f(\overline{\underline{\phi}}_k,\overline{\rho}_k) = \ell n \overline{\rho}_k + 1$

and the decision statistic is

$$h_k'(Z_N) = N(\ell n \rho_k) + n_k \ell n \ N \qquad (3.24)$$

which is the statistic reported in [3] for UAR model fitting.

In the next section we prove the consistency of the decision rule.

4. Consistency of the decision rule

Definition: Let $P_j(Z_N|C_i)$ denote the probability that we choose the model C_j when the true model is C_i. Let the observations obey the model C_i. The decision rule is said to be consistent if $P_j(Z_N|C_i) \to 0$ uniformly as $N \to \infty$ for all $j \neq i$.

For simplicity we assume that there are two hypotheses C_1 and C_2. Let the observations obey the hypothesis C_1. Then C_2 could belong to one of the following two cases:

Case (i): (Over-specified hypothesis)

C_2 is overspecified w.r.t C_1 if there exists a $\underset{\sim}{\phi}' \in R^{n_k}$ such that

$$A_2(\underset{\sim}{\phi}',D) = A_1(\underset{\sim}{\phi},D) \text{ and } B_2(\underset{\sim}{\phi}',D) = B_1(\underset{\sim}{\phi},D)$$

for $\underset{\sim}{\phi}^T = (\phi_1, \phi_2, \ldots, \phi_{n_1})$, $\phi_j \neq 0$, $j=1,2,\ldots,n_1$

Example: Consider the hypotheses C_1, C_2, and C_3 defined by the equations E_1, E_2, and E_3:

E_1: $(1 + \phi_1 D + \phi_2 D^{-1} + \phi_3 D^{-2})y(t) = \omega(t)$

E_2: $(1 + \phi_1 D + \phi_2 D^2 + \phi_3 D^{-1} + \phi_4 D^{-2})y(t) = \omega(t)$

E_3: $(1 + \phi_1 D^{-1} + \phi_2 D^{-2} + \phi_3 D^{-3})y(t) = \omega(t)$

C_2 is overspecified w.r.t C_1 but C_3 is not overspecified w.r.t H_1.

Case (ii): All other models not covered by Case (i)

We state and prove a theorem which establishes the consistency of the decision rule.

<u>Theorem 4</u>:   The decision rule given in eq. (3.19) is consistent.

For the models covered by case (i) $P_2(Z_N|C_1) \leq O(K_1/(\ell nN)^2)$, $k_1 > 0$
and for the models covered by case (ii),

$P_2(Z_N|C_1) \leq O(K_2/N)$, $k_2 > 0$.

This theorem is proved in Appendix III.

5. Discussion

The decision rule developed here completely solves the problem of the choice of neighbors for one-dimensional AR models for a given empirical series. The decision rules developed in [1],[2][3] have covered only the UAR models. Only Whittle [4] has considered the problem of BAR models in connection with a line transect, but no proof is given for the consistency of his decision rule.

The hypotheses C_i, $1 \leq i \leq r$ defined here include both unilateral and bilateral AR models. The decision rule is consistent, transitive, and yields a quantitative explanation for the principle of parsimony used in model building. The asymptotic analysis given here holds for large values of N, about 100-200. The Bayesian approach has two advantages: (1) It yields consistent decision rules for choosing the correct model; (2) the analysis yields an explicit expression for the probability density of transforms of observations given the model that the observations obey. This expression could be used for classification purposes.

6. Applications and extensions

Assume that we are given a sequence of sampled gray levels along a row of an image.   We can consider UAR models of orders one and two and BAR models of orders one and two as a set of plausible models and use the theory developed here to choose the best model with minimum probability of error.

As the Bayesian approach yields an explicit expression for the probability density of observations given a model, better rules can be developed for classification purposes.

The theory can be easily extended to cover bilateral auto-repressive and moving average (BARMA) models by appropriately modifying the structure of the transfer function and the assoc-iated stability conditions.

The extension of the theory described here to stationary random fields is considered in Part II.

Part II

Statistical Inference Theory
Applied to Texture Representation

1. Introduction

We are interested in developing statistical models for textures. In particular, we are interested in applying the theory of statistical inference of stationary random fields to images. We assume that the textures under consideration are sample functions of stationary random fields, not necessarily isotropic. The organization of the paper is as follows: In the rest of this section we review earlier research done in image modeling and motivate the inference approach. In Section 2 we formulate the problem and develop a decision rule for inferring the correct model with minimum probability of error. We also assert the consistency of the decision rule. Section 3 compares the theory developed here with other known approaches in the literature. Applications and possible extensions are indicated in Section 4.

1.1 Types of models

Early attempts at image modeling [14],[15] applied one-dimensional time series analysis to two-dimensional images. By concatenation of successive rows, a one-dimensional time series is generated. Seasonal autoregressive integrated moving average models [14] and seasonal autoregressive models [15] have been fitted to this time series.

It is intuitively clear that any model should reflect the two-dimensional nature of the image. Tou et al. [16] considered two-dimensional autoregressive and moving average models for textures. By differencing along rows and/or columns nonstationarity in the series is removed. By inspection of autocorrelation functions tentative models are determined and the parameters are estimated by least square methods. In a subsequent paper [17] classification rules are given based on differences between the gray levels of the original texture and regenerated texture.

Recently, Kashyap [18] has suggested a two-dimensional unilateral autoregressive model for images. A vector of sufficient statistics has been derived for the parameters of the model which by definition possess all the information in the samples. Consistent rules have been developed to determine the width of the one-sided neighborhood.

It is not clear if these unilateral models, though they are two-dimensional, are appropriate for images, since for images

the neighborhood dependence extends in all directions. Though
Whittle [4] has considered equivalent unilateral schemes for
a given bilateral scheme, there are instances where there exists
no equivalent finice unilateral autoregressive model (UAR) for
a given bilateral autoregressive model [4].

Pratt [19][20] and Gagalowicz [21] have suggested two-
dimensional spatial filter models which transform a sequence of
white noise variables into the observed structure. No attempts
have been made to use inference methods for identifying the
models.

Stochastic partial differential equations [22] have also been
suggested as models for images. The discretized equivalents of
partial differential equations have been fitted using least
square techniques. These models cover nonseparable, nearly iso-
tropic images. No attempts have been made to infer the models
from the data.

Two-dimensional linear estimation techniques have been con-
sidered for textures [23]. The gray level at the element (i,j)
is assumed to depend on neighbors within a window surrounding the
(i,j) element. However, the dimensions of the window are deter-
mined by using Akaike's statistics, which apply to a singly
indexed sequence. A mean square criterion is used for determin-
ing the coefficients of the model.

By operating on an array of independent

and identically distributed random variables, using a set of independent parameters which control the directionality and graininess of the random field generated, textures are synthesized in [24][25]. Results are given for binary first order isotropic Markov random fields[24]. The number of parameters is proportional to the square of the number of gray levels in the texture and may be large for real textures.

There are many shortcomings in the models that have been discussed above. Some of the models view an image as a concatenation of rows, which is clearly inadequate. Some consider two-dimensional but unilateral models. Even in these cases, except for [18], no attempts have been made to infer the order of dependency or dimension of neighborhood. Though two-dimensional neighbhorhood models have been considered in [19][20][21], no attempts have been made to theoretically justify the order of the models. The work of Pratt and Gagalowicz is motivated by experimental results reported earlier by Julesz. The models are built with the idea of matching the correlations of the parent texture and this may not always yield consistent models.

1.2 Inference of models

We believe that any realistic model should be inferred from the data by using statistical inference theory. This is a powerful approach since the underlying probability distribution of a texture can be inferred in a consistent way by using the tools of system identification.

Not much work has been done in the area of image modeling using statistical inference of random fields. This involves assertions about the probability distribution of the observed data. This is usually accomplished by considering parametric forms of probability distributions with a finite number of parameters which is reasonably smaller than the number of observations. Statistical inference is then concerned with choosing among the various parametric descriptions of the underlying data. As we are interested in building models for images, our basic models will include bilateral models. We will also include some unilateral models to check if the bilateral models are preferred to unilateral models, for images.

For the reasons given in Part I, we will do the analysis in the spectral domain and take a Bayesian approach to fitting models.

2. Decision rule

The case of a random field is exactly analogous to the one-dimensional case that we have already considered. In this section, we give the equation for the random field and state the corresponding stability conditions. We state theorems parallel to the ones in Section 3 and suggest the decision rule that chooses the correct model with minimum probability of error.

We are given a set of observations $y(\underset{\sim}{s})$, $\underset{\sim}{s} \in \Omega_S$, $\underset{\sim}{s} = (s_1, s_2)^T$, from Ω_S, a grid of dimension $N_1 \times N_2$ and $1 \leq s_i \leq N_i$, $i = 1, 2$, from a stationary random field and r mutually exclusive compound hypotheses C_i, $1 \leq i \leq r$. We define the ith parametric form of the random field as follows:

$$E_i: \quad \sum_{\underset{\sim}{q} \in Q} A(\underset{\sim}{q}) \, y(\underset{\sim}{s} + \underset{\sim}{q}) = u(\underset{\sim}{s}) \tag{2.1}$$

where Q is a finite set of two-dimensional vector shifts and $u(\underset{\sim}{s})$ is an independent and identically distributed Gaussian random field with mean zero and variance ρ. $(A(\underset{\sim}{q}), \underset{\sim}{q} \in Q)^T$ and ρ are unknown.

Let $\underset{\sim}{\phi}^T = (A(\underset{\sim}{q}), \underset{\sim}{q} \in Q)$ and $\underset{\sim}{\theta}^T = (\underset{\sim}{\phi}^T, \rho)$

Define the shift operator,

$$\underset{\sim}{D}^{\underset{\sim}{q}} y(\underset{\sim}{s}) = y(\underset{\sim}{s} + \underset{\sim}{q}), \quad \underset{\sim}{D}^T = (D_1, D_2)$$

$$\underset{\sim}{D}^{\underset{\sim}{q}} = D_1^{q_1} D_2^{q_2}$$

such that

$$D_1^{q_1} D_2^{q_2} y(\underset{\sim}{s}) = y(s_1 + q_1, s_2 + q_2)$$

and

$$H_i(\underset{\sim}{D},\underset{\sim}{\phi}) = \sum_{\underset{\sim}{q} \in Q} A_i(\underset{\sim}{q})\underset{\sim}{D}^{\underset{\sim}{q}}$$

Then (2.1) can be represented as

$$H_i(\underset{\sim}{D},\underset{\sim}{\phi})y(\underset{\sim}{s}) = u(\underset{\sim}{s}) \tag{2.2}$$

Consider the $\underset{\sim}{z}$ transform of (2.2) where $\underset{\sim}{z}^T = (z_1,z_2)$ is a complex vector

$$H_i(\underset{\sim}{z},\underset{\sim}{\phi})Y(\underset{\sim}{z}) = U(\underset{\sim}{z}) \tag{2.3}$$

We make the following assumption (A1) about the stability of the equation (2.1).

A1): $H_i(\underset{\sim}{z},\underset{\sim}{\phi}) \neq 0$ for $|z_1|=|z_2|= 1$.

Let $(x(\underset{\sim}{\lambda}), \underset{\sim}{\lambda}\in\Omega_\lambda)$, $\underset{\sim}{\lambda}= (\lambda_1,\lambda_2)^T$,

$\lambda_i = 2\pi k_i/N$, $0 \leq k_i < N_i$, $i=1,2, N=N_1N_2$

denote the finite Fourier transforms of the observations $(y(\underset{\sim}{s}), \underset{\sim}{s}\in\Omega_S)$ from the random field. As discussed in Section 3, Part 1, the decision rule that chooses the correct model with minimum probability of error is:

Decide hypothesis is k^* if

$$k^* = \arg \max_{k}\{P(C_i|x(\underset{\sim}{\lambda}), \underset{\sim}{\lambda}\in\Omega_\lambda\} \tag{2.4}$$

We have

$$P(C_i|x(\underset{\sim}{\lambda}), \underset{\sim}{\lambda}\in\Omega_\lambda) = \frac{p(x(\underset{\sim}{\lambda}), \underset{\sim}{\lambda}\in\Omega_\lambda|C_i)P(C_i)}{\sum_{j=1}^{r} p(x(\underset{\sim}{\lambda}),\underset{\sim}{\lambda}\in\Omega_\lambda|C_j)P(C_j)} \tag{2.5}$$

and

$$p(x(\underset{\sim}{\lambda}), \underset{\sim}{\lambda}\in\Omega_\lambda|C_i) = \int p(x(\underset{\sim}{\lambda}),\underset{\sim}{\lambda}\in\Omega_\lambda|\underset{\sim}{\theta},C_i)p(\underset{\sim}{\theta}|C_i)d\underset{\sim}{\theta} \tag{2.6}$$

We now state Theorem 1' which is a generalization of Theorem 1.

Theorem 1':

Let the observations ($y(\underset{\sim}{s})$, $\underset{\sim}{s} \in \Omega_S$) obey the kth equation E_k. Then as the rectangle Ω_S becomes large in all dimensions of $\underset{\sim}{s}$, the finite Fourier transform is approximately distributed normally with mean zero and independently at different frequencies with variances

$$S_{yk}(e^{i\underset{\sim}{\lambda}^T}, \underset{\sim}{\theta}) = \rho H_k(e^{j\underset{\sim}{\lambda}^T}, \underset{\sim}{\phi}) \ H_k^*(e^{j\underset{\sim}{\lambda}^T}, \underset{\sim}{\phi}) \qquad \text{for all } \underset{\sim}{\lambda} \in \Omega_\lambda$$

We need the following assumption (A2'):

A2') The first and second derivatives of $\underset{\underset{\sim}{\lambda} \in \Omega_\lambda}{\Sigma} \ell n ||H_k(e^{j\underset{\sim}{\lambda}^T}, \underset{\sim}{\phi})||^2$ and $\underset{\underset{\sim}{\lambda} \in \Omega_\lambda}{\Sigma} ||\underset{\sim}{y}(\underset{\sim}{\lambda})||^2 / ||H_k(e^{j\underset{\sim}{\lambda}^T}, \underset{\sim}{\phi})||^2$ w.r.t. $\underset{\sim}{\phi}$ exist for all $\underset{\sim}{\phi} \in R^{n_k}$

We state Theorem 2', as a generalization of Theorem 2.

Theorem 2': As the rectangle Ω_S becomes large in all dimensions of $\underset{\sim}{s}$,

$$\ell n \ p(x(\underset{\sim}{\lambda}), \ \underset{\sim}{\lambda} \in \Omega_\lambda, \underset{\sim}{\theta} | C_k)$$

$$= -(N/2) f(\overline{\underset{\sim}{\theta}}_k) - (N/2) ((\rho - \overline{\rho}_k)^2 \overline{d}_k + (\underset{\sim}{\phi} - \overline{\underset{\sim}{\phi}}_k)^T \overline{\underset{\sim}{Q}}_k (\underset{\sim}{\phi} - \overline{\underset{\sim}{\phi}}_k)$$

$$+ (\rho - \overline{\rho}_k)(\underset{\sim}{\phi} - \overline{\underset{\sim}{\phi}}_k)^T \overline{\underset{\sim}{s}}_k + O(||\underset{\sim}{\theta} - \overline{\underset{\sim}{\theta}}_k||^3) \tag{2.7}$$

where

$$f(\overline{\underset{\sim}{\theta}}_k) = 1 + \ell n \overline{\rho}_k + (1/N) \underset{\underset{\sim}{\lambda} \in \Omega_\lambda}{\Sigma} \ell n || H_k(e^{j\underset{\sim}{\lambda}^T}, \overline{\underset{\sim}{\phi}}_k)||^2 \tag{2.8}$$

$$\overline{\rho}_k = \frac{1}{N} \underset{\underset{\sim}{\lambda} \in \Omega_\lambda}{\Sigma} \cdot ||x(\underset{\sim}{\lambda})||^2 / ||H_k(e^{j\underset{\sim}{\lambda}^T}, \overline{\underset{\sim}{\phi}}_k)||^2 \tag{2.9}$$

$$\overline{\phi}_k = \min_{\underset{\sim}{\phi}} \{ \frac{1}{N} \sum_{\underset{\sim}{\lambda} \in \Omega_\lambda} \ell n ||H_k(e^{j\underset{\sim}{\lambda}^T}, \overline{\underset{\sim}{\phi}})||^2$$

$$+ \ell n(\frac{1}{N} \sum_{\underset{\sim}{\lambda} \in \Omega_\lambda} ||x(\underset{\sim}{\lambda})||^2 / ||H_k(e^{j\underset{\sim}{\lambda}^T}, \underset{\sim}{\phi})||^2) \tag{2.10}$$

$$\overline{d}_k = \partial^2 f(\underset{\sim}{\phi}, \rho)/\partial \rho^2 \Big|_{\substack{\underset{\sim}{\phi}=\overline{\underset{\sim}{\phi}}_k \\ \rho=\overline{\rho}_k}} \qquad \text{(a scalar)} \tag{2.11}$$

$$\overline{Q}_k = \frac{1}{2} \left. \frac{\partial^2 f(\underset{\sim}{\phi}, \rho)}{\partial \phi_i \partial \phi_j} \right|_{\substack{\underset{\sim}{\phi}=\overline{\underset{\sim}{\phi}}_k \\ \rho=\overline{\rho}_k}} \qquad \begin{array}{l} (n_k \times n_k \text{ matrix}) \\ \\ 1 \leq i, j \leq n_k \end{array} \tag{2.12}$$

$$\overline{S}_k = \partial^2 f(\underset{\sim}{\phi}, \rho)/\partial \rho \partial \phi_i \Big|_{\substack{\underset{\sim}{\phi}=\overline{\underset{\sim}{\phi}}_k \\ \rho=\overline{\rho}_k}} \qquad (n_k \times 1) \text{ vector} \tag{2.13}$$

and

$$f(\underset{\sim}{\phi}, \rho) = \ell n \rho + (1/N) \sum_{\underset{\sim}{\lambda} \in \Omega_\lambda} \ell n ||H_k(e^{j\underset{\sim}{\lambda}^T}, \underset{\sim}{\phi})||^2$$

$$+ \frac{1}{N\rho} \sum_{\underset{\sim}{\lambda} \in \Omega_\lambda} ||x(\underset{\sim}{\lambda})||^2 / ||H_k(e^{j\underset{\sim}{\lambda}^T}, \underset{\sim}{\phi})||^2 \tag{2.14}$$

Integrating $p(x(\lambda), \lambda \in \Omega_\lambda | \underset{\sim}{\theta}, C_k)$ over the prior probability density $p(\underset{\sim}{\theta} | C_k)$, $1 \leq k \leq r$ by using the aysmptotic theory of integration, we derive an expression for the posterior probability density, $P(C_k | x(\underset{\sim}{\lambda}), \underset{\sim}{\lambda} \in \Omega_\lambda)$ given in Theorem 3'.

Theorem 3': As the rectangle Ω_S becomes large in all dimensions of $\underset{\sim}{s}$,

$$\ln P(C_k | x(\underset{\sim}{\lambda}), \underset{\sim}{\lambda} \in \Omega_\lambda)$$

$$= -(N/2) f(\overline{\underset{\sim}{\theta}}_k) + \ln p(\overline{\underset{\sim}{\theta}} | C_k) + \frac{1}{2}(n_k+1)\ln(2\pi/N)$$

$$+ (N/2)\ln 2\pi - \frac{1}{2}\ln(\det F_{n_k}(-g(\underset{\sim}{\theta};N)))_{\underset{\sim}{\theta} = \overline{\underset{\sim}{\theta}}_k}$$

$$+ \ln P(C_k) - \ln p(x(\underset{\sim}{\lambda}), \underset{\sim}{\lambda} \in \Omega_\lambda) \tag{2.15}$$

$$g(\underset{\sim}{\theta};N) = -[(\rho - \overline{\rho}_k)^2 \overline{d}_k + (\underset{\sim}{\phi} - \overline{\underset{\sim}{\phi}}_k)^T \underset{\sim}{Q}_k (\underset{\sim}{\phi} - \overline{\underset{\sim}{\phi}}_k)$$

$$+ (\rho - \overline{\rho}_k)(\underset{\sim}{\phi} - \overline{\underset{\sim}{\phi}}_k)^T \underset{\sim}{S}_k]$$

The proof given in Appendix II can be easily extended to prove this theorem, and the comments following Theorem 3 hold here also.

Now we give the approximate decision rule that chooses the correct model with minimum probability of error:

Decide hypothesis is k* if

$$k^* = \arg \max_k \{h_k(x(\underset{\sim}{\lambda},, \underset{\sim}{\lambda} \in \Omega_\lambda)\} \tag{2.17}$$

where

$$h_k(x(\underset{\sim}{\lambda}), \underset{\sim}{\lambda} \in \Omega_\lambda) = -(N/2) f(\overline{\underset{\sim}{\theta}}_k)$$

$$-(n_k/2)\ln N + \ln p(\overline{\underset{\sim}{\theta}}_k | C_k) - \frac{1}{2}\ln(\det F_{n_k}(-g(\underset{\sim}{\theta};N)) \Big|_{\underset{\sim}{\theta} = \overline{\underset{\sim}{\theta}}_k} \tag{2.18}$$

For practical applications, we suggest a simplified decision rule:

Decide hypothesis is k* if

$$k^* = \arg \min_k \{h_k'(x(\lambda), \underset{\sim}{\lambda} \in \Omega_\lambda\} \tag{2.19}$$

where

$$h'_k(x(\lambda), \lambda \in \Omega_\lambda) = N \ f(\overline{\theta}_k) + n_k \ell nN \qquad (2.20)$$

The decision rules given in (2.17) and (2.19) are consistent. A proof similar to that in Appendix III can be given to establish their consistency.

3. Discussion and comparisons

The topic of statistical inference on random fields, which is of primary interest in this paper, has been previously considered by Whittle [4] and recently by Larimore[9]. Whittle has developed spectral methods for stationary autoregressive scalar random fields. But this was before the development of algorithms for fast Fourier transforms and no attempt has been made to prove that the criterion of choosing the right model is consistent.

Larimore has extended Whittle's method for the case of vector random fields. But Akaike's criterion has been used for choosing the best model. As observed in [5], there exists no proof for the optimality of Akaike's rule and recently it has been proved that Akaike's rule is inconsistent [26].

We have suggested a consistent decision rule that chooses a correct model with minimum probability of error. The theory can be extended to include moving average terms in the stochastic difference equation. This modifies the conditions for stability and the numerical computations for estimating the coefficients become more complex.

We believe that this approach will be of use in image modeling. So far, researchers in image modeling have either considered unilateral models or have not used system identification methods to choose the correct model. The inference procedure developed here chooses the correct model with minimum probability of error

for unilateral as well as bilateral models.

This approach yields an explicit expression for the probability density of transforms of observations given the model the observations obey. This has not been done before for bilateral models. The expression for the probability density of transforms of observations could be used for classification purposes. This approach should result in good classification strategies for textures.

It should be pointed out that the theory developed here is based on the assumption that the random field is Gaussian. This assumption has often been used in the literature on image modeling [15][17][18].

Appendix I

We prove Theorem 2.

Consider equation (3.10) repeated below:

$$\ell np(z(\lambda_1), z(\lambda_2), \ldots z(\lambda_N) | \underset{\sim}{\phi}, \rho, C_k)$$

$$= -\frac{N}{2}\ell n 2\pi\rho - \frac{1}{2}\sum_{i=1}^{N} \ell n ||H_k(e^{j\lambda_i}, \underset{\sim}{\phi})||^2$$

$$- \frac{1}{2\rho}\sum_{i=1}^{N} ||z(\lambda_i)||^2 / ||H_k(e^{j\lambda_i}, \underset{\sim}{\phi})||^2 \qquad (1)$$

We first compute the maximum likelihood estimates of ϕ and ρ under the hypothesis C_k.

Differentiating (1) w.r.t. ρ and equating to zero,

$$\tilde{\rho}_k(\underset{\sim}{\phi}) = \frac{1}{N}\sum_{i=1}^{N} ||z(\lambda_i)||^2 / ||H_k(e^{j\lambda_i}, \underset{\sim}{\phi})||^2 \qquad (2)$$

Substituting (2) in (1), the maximum likelihood estimate (m.l.e) $\overline{\underset{\sim}{\phi}}_k$ is given by

$$(3)$$

$$\overline{\underset{\sim}{\phi}}_k = \min_{\underset{\sim}{\phi}} \{ \frac{1}{N}\sum_{i=1}^{N} \ell n ||H_k(e^{j\lambda_i}, \underset{\sim}{\phi})||^2 + \ell n [(\frac{1}{N})\sum_{i=1}^{N} ||z(\lambda_i)||^2 / ||H_k(e^{j\lambda_i}, \underset{\sim}{\phi})||^2] \}$$

and

$$\overline{\rho}_k = \tilde{\rho}_k(\overline{\underset{\sim}{\phi}}_k) = \frac{1}{N}\sum_{i=1}^{N} ||z(\lambda_i)||^2 / ||H_k(e^{j\lambda_i}, \overline{\underset{\sim}{\phi}}_k)||^2 \qquad (4)$$

Rewrite (1) as follows:

$$\ell np(z(\lambda_1), z(\lambda_2), \ldots z(\lambda_N) | \underset{\sim}{\phi}, \rho, C_k)$$

$$= -\frac{N}{2}\ell n 2\pi - \frac{N}{2} f(\underset{\sim}{\phi}, \rho) \qquad (5)$$

4. Extensions

We propose to use the models developed here to develop better classification rules for textures. The theory developed here can be extended to the case of vector random fields. This will be useful to build models for observation pairs such as:

a) (gray level, edge value)

b) (gray level, average gray level over a neighborhood)

We will treat these extensions in subsequent papers.

where

$$f(\underline{\phi},\rho) = \ell n\rho + \frac{1}{N}\sum_{i=1}^{N}\ell n||H_k(e^{j\lambda_i},\underline{\phi})||^2$$

$$+(1/N\rho)\sum_{i=1}^{N}||z(\lambda_i)||^2/||H_k(e^{j\lambda_i},\underline{\phi})||^2 \qquad (6)$$

Expanding $f(\underline{\phi},\rho)$ as a Taylor series in $\underline{\phi}$ and ρ at $\underline{\phi}=\overline{\underline{\phi}}_k$ and $\rho=\overline{\rho}_k$, we have

$$f(\underline{\phi},\rho) = f(\overline{\underline{\phi}}_k,\overline{\rho}_k) + \left.\frac{\partial f(\underline{\phi},\rho)}{\partial\rho}\right|_{\substack{\rho=\overline{\rho}_k \\ \underline{\phi}=\overline{\underline{\phi}}_k}} \cdot (\rho-\overline{\rho}_k)$$

$$+ \left.\frac{\partial f(\underline{\phi},\rho)}{\partial\underline{\phi}}\right|_{\substack{\underline{\phi}=\overline{\underline{\phi}}_k \\ \rho=\overline{\rho}_k}}^T \cdot (\underline{\phi}-\overline{\underline{\phi}}_k) + \frac{1}{2}(\rho-\overline{\rho}_k)^2 \left.\frac{\partial^2 f(\underline{\phi},\rho)}{\partial\rho^2}\right|_{\substack{\underline{\phi}=\overline{\underline{\phi}}_k \\ \rho=\overline{\rho}_k}}$$

$$+ \frac{1}{2}(\underline{\phi}-\overline{\underline{\phi}}_k)^T \left.\frac{\partial^2 f(\underline{\phi},\rho)}{\partial\phi_i\partial\phi_j}\right|_{\substack{\underline{\phi}=\overline{\underline{\phi}}_k \\ \rho=\overline{\rho}_k}} \cdot (\underline{\phi}-\overline{\underline{\phi}}_k) + \qquad 1\in i,j\leq n_k \quad (7)$$

$$(\rho-\overline{\rho}_k)(\underline{\phi}-\overline{\underline{\phi}}_k)^T \left.\frac{\partial^2 f(\underline{\phi},\rho)}{\partial\rho\partial\phi_i}\right|_{\substack{\underline{\phi}=\overline{\underline{\phi}}_k \\ \rho=\overline{\rho}_i}} + O(||\underline{\phi}-\overline{\underline{\phi}}_k||^3)$$

By definition of $\overline{\rho}_k$ and $\overline{\underline{\phi}}_k$,

$$\left.\frac{\partial f(\underline{\phi},\rho)}{\partial\rho}\right|_{\rho=\overline{\rho}_k} \text{ and } \left.\frac{\partial f(\underline{\phi},\rho)}{\partial\underline{\phi}}\right|_{\substack{\underline{\phi}=\overline{\underline{\phi}}_k \\ \rho=\overline{\rho}_k}} \text{ are zero,}$$

yielding

$$f(\underline{\phi},\rho) = f(\overline{\underline{\phi}}_k,\overline{\rho}_k) + \frac{1}{2}(\rho-\overline{\rho}_k)^2 \left.\frac{\partial^2 f(\underline{\phi},\rho)}{\partial\rho^2}\right|_{\substack{\underline{\phi}=\overline{\underline{\phi}}_k \\ \rho=\overline{\rho}_k}}$$

$$+ \frac{1}{2}(\underline{\phi}-\overline{\underline{\phi}}_k)^T \left.\frac{\partial^2 f(\underline{\phi},\rho)}{\partial\phi_i\,\partial\phi_j}\right|_{\substack{\underline{\phi}=\overline{\underline{\phi}}_k \\ \rho=\overline{\rho}_k}} (\underline{\phi}-\overline{\underline{\phi}}_k) \;+\; \qquad 1\le i,j\le n_k$$

$$(\rho-\overline{\rho}_k)(\underline{\phi}-\overline{\underline{\phi}}_k)^T \left.\frac{\partial^2 f(\underline{\phi},\rho)}{\partial\rho\partial\underline{\phi}}\right|_{\substack{\underline{\phi}=\overline{\underline{\phi}}_k \\ \rho=\overline{\rho}_k}} + O(||\underline{\phi}-\overline{\underline{\phi}}_k||^3) \qquad (8)$$

Using (8) and (3.13--3.15) in (5), we have

$$\ell np(z(\lambda_1),z(\lambda_2),\ldots z(\lambda_N)|\underline{\phi},\rho,C_k)$$

$$= -\frac{N}{2}\ell n2\pi - \frac{N}{2}(f(\overline{\underline{\phi}}_k,\overline{\rho}_k) + (\rho-\overline{\rho}_k)^2\overline{d}_k +$$

$$(\underline{\phi}-\overline{\underline{\phi}}_k)^T\overline{Q}_k(\underline{\phi}-\overline{\underline{\phi}}_k) + (\rho-\overline{\rho}_k)(\underline{\phi}-\overline{\underline{\phi}}_k)^T\overline{S}_k + O(||\underline{\phi}-\overline{\underline{\phi}}_k||^3) \qquad (9)$$

Q.E.D.

Appendix II

We prove Theorem 3. We first state a lemma [10] to be used in the proof of the theorem.

<u>Lemma 1</u>: Consider the integral

$$G(N) = \int_R h(\underset{\sim}{\theta})\ \exp[g(\underset{\sim}{\theta},N)]d\underset{\sim}{\theta} \tag{10}$$

where

R is an n-dimensional domain in the Euclidean n-space

$\underset{\sim}{\theta}$ is an n-dimensional vector

N is a large positive integer

$g(\underset{\sim}{\theta},N)$ (a) is a bounded function for N large and assumes

an absolute maximum at an interior point

$$\overline{\underset{\sim}{\theta}}(N) = (\overline{\underset{\sim}{\theta}}_1(N),\dots,\overline{\underset{\sim}{\theta}}_n(N))^T$$

(b) $F_n(-g(\overline{\underset{\sim}{\theta}},N)) > c > 0$ hold in R for N large

and

$$F_n(-g(\underset{\sim}{\theta},N)) = \det||-g_{\theta_i \theta_j}(\underset{\sim}{\theta},N)|| \qquad\qquad 1 \le i,j \le n$$

Here $g_{\theta_i \theta_j}(\underset{\sim}{\theta},N)$ is the second order partial derivative of $g(\theta,N)$ with respect to θ_i and θ_j.

Then

$$G(N) = (\frac{2\pi}{N})^{n/2}\ \frac{[\exp(g(\overline{\underset{\sim}{\theta}}(N),N)]^N h(\overline{\underset{\sim}{\theta}}(N))}{[F_n(-g(\underset{\sim}{\theta},N))]^{1/2}}\Bigg|_{\underset{\sim}{\theta}=\overline{\underset{\sim}{\theta}}(N)}$$

Proof of Theorem 3:

We have to perform the integration

$$\iint p(z(\lambda_1),\dots,z(\lambda_N)|\underset{\sim}{\phi},\rho,C_k)p(\underset{\sim}{\phi},\rho|C_k)d\underset{\sim}{\phi}d\rho \tag{12}$$

Substituting for $p(z(\lambda_1),\ldots,z(\lambda_N)|\underset{\sim}{\phi},\rho,C_k)$ from (9), we get

LHS of (12)

$$= (1/2\pi)^{N/2}\exp(-\frac{N}{2}f(\overline{\underset{\sim}{\phi}}_k,\overline{\rho}_k))\int\exp\{\frac{N}{2}(-(\rho-\overline{\rho}_k)^2\overline{d}_k-(\underset{\sim}{\phi}-\overline{\underset{\sim}{\phi}}_k)^T\overline{\underset{\sim}{Q}}_k(\underset{\sim}{\phi}-\overline{\underset{\sim}{\phi}}_k)$$

$$-\ (\rho-\overline{\rho}_k)(\underset{\sim}{\phi}-\overline{\underset{\sim}{\phi}}_k)^T\overline{\underset{\sim}{S}}_k\}p(\underset{\sim}{\phi},\rho|C_k)d\underset{\sim}{\phi}d\rho \tag{13}$$

Identifying the terms in (13) with the terms defined in Lemma 1, we get

$$p(Z_N|C_k) = \frac{(1/2\pi)^{N/2}\exp(-\frac{N}{2}f(\overline{\underset{\sim}{\phi}}_k,\overline{\rho}_k))(\frac{2\pi}{N})^{(n_k+1)/2}p(\overline{\underset{\sim}{\phi}}_k,\overline{\rho}_k|C_k)}{[F_k(-g(\underset{\sim}{\theta},N))]^{1/2}}\Bigg|_{\underset{\sim}{\theta}=\overline{\underset{\sim}{\theta}}(N)} \tag{14}$$

where

$$g(\underset{\sim}{\theta},N) = -[(\rho-\overline{\rho}_k)^2\overline{d}_k + (\underset{\sim}{\phi}-\overline{\underset{\sim}{\phi}}_k)^T\overline{\underset{\sim}{Q}}_k(\underset{\sim}{\phi}-\overline{\underset{\sim}{\phi}}_k)$$

$$+ (\rho-\overline{\rho}_k)(\underset{\sim}{\phi}-\overline{\underset{\sim}{\phi}}_k)^T\overline{\underset{\sim}{S}}_k] \tag{15}$$

Appendix III

We prove the consistency of the decision rule given in Theorem 3.

We have

$$P_2(Z_N|C_1) = \text{Prob}[h_1 > h_2 | C_1]$$

We evaluate $\text{Prob}[h_1 > h_2 | C_1]$ for the two cases mentioned in Section 4.

Case (i):

$$\overline{\rho}_2 \leq \overline{\rho}_1$$

Now,

$$h_1 - h_2 = N(\ln \overline{\rho}_1 - \ln \overline{\rho}_2) + \sum_{i=1}^{N} \ln ||H_1(e^{j\lambda_i}, \overline{\phi}_1)||^2$$

$$- \sum_{i=1}^{N} \ln ||H_2(e^{j\lambda_i}, \overline{\phi}_2)||^2 + (n_1 - n_2)\ln N + \xi_1 \qquad (18)$$

where

$$\xi_1 = 2\ln[p(\overline{\phi}_2, \overline{\rho}_2|C_2)/p(\overline{\phi}_1, \overline{\rho}_1|C_1)]$$

$$+ \ln \det F_1(-g(\underset{\sim}{\theta}, N)) \Big|_{\underset{\sim}{\theta} = \overline{\underset{\sim}{\theta}}(N)}$$

$$- \ln \det F_2(-g(\underset{\sim}{\theta}, N)) \Big|_{\underset{\sim}{\theta} = \overline{\underset{\sim}{\theta}}(N)}$$

$$+ 2\ln[P(C_2)/P(C_1)] \qquad (19)$$

Using $\ln x \leq x - 1$,

$$h_1 - h_2 \leq N(\overline{\rho}_1 - \overline{\rho}_2) + \sum_{i=1}^{N} \ln ||H_1(e^{j\lambda_i}, \overline{\phi}_1)||^2$$

$$- \sum_{i=1} \ln ||H_2(e^{j\lambda_i}, \overline{\phi}_2)||^2 + (n_1 - n_2)\ln N + \xi_1$$

$$\frac{h_1 - h_2}{n_2 - n_1} \leq \frac{N(\overline{\rho}_1 - \overline{\rho}_2)}{(n_2 - n_1)\overline{\rho}_1} + \frac{1}{(n_2 - n_1)} \, .$$

$$\left[\sum_{i=1}^{N} \ell n ||H_1(e^{j\lambda_i},\overline{\phi}_1)||^2 - \sum_{i=1}^{N} \ell n ||H_2(e^{j\lambda_i},\overline{\phi}_2)||^2\right] - \ell nN + \frac{\xi_1}{(n_2-n_1)}$$

Recall $\quad \overline{\rho}_1 = \frac{1}{N} \sum_{i=1}^{N} \frac{||z(\lambda_i)||^2}{H_1(e^{j\lambda_i},\overline{\phi}_1)||^2}$ (19)

From Theorem 1, $z(\lambda_i)$ is a complex normal random variable with mean zero and variance $S_{y1}(e^{j\lambda_i},\phi,\rho)$ and hence $||z(\lambda_i)||^2$ is distributed as an exponential variate with mean $S_{y1}(e^{j\lambda_i},\phi,\rho)$. Equivalently, $||z(\lambda_i)||^2$ is distributed as $\frac{1}{2}S_{y1}(e^{j\lambda_1},\phi,\rho)\chi^2(2)$, where $\chi^2(2)$ is a chi-square distribution with two degrees of freedom. Also from Theorem 1, $z(\lambda_i)$ is independent of $z(\lambda_j)$ for $i \neq j$ and hence (19) represents a sum or weighted independent chi-squared variables. The standard method [27] of obtaining the distribution of sum of independent weighted chi-squareds is to approximate it by a multiple, $k\chi^2(v)$, of a chi-squared variable whose mean and degrees of freedom are determined by equating first and second order moments. For our purposes, it suffices to mention that $\overline{\rho}_1$ is a known multiple chi-squared random variable.

Let $\eta_1 = N(\overline{\rho}_1-\overline{\rho}_2)/\overline{\rho}_1 + \frac{1}{(n_2-n_1)}$.

$$\left[\sum_{i=1}^{N} \ell n ||H_1(e^{j\lambda_i},\overline{\phi}_1)||^2 - \sum_{i=1}^{N} \ell n ||H_2(e^{j\lambda_i},\overline{\phi}_2)||^2\right]$$

$+\xi_1/(n_2-n_1)$ be a random variable with finite mean and variance σ_1^2 .

Then

$$P_2(Z_N|C_1) = \text{Prob}[h_1 > h_2|C_1]$$

$$\leq \text{Prob}[\eta_1 > \ell nN] \leq O(K_1/(\ell nN)^2) \qquad K_1 > 0 \qquad (20)$$

by the Chebychev inequality.

Case (ii):

$$\bar{\rho}_2 > \bar{\rho}_1 \qquad (21)$$

Since $\bar{\rho}_1$ is a maximum likelihood estimate of ρ, we have

$$\bar{\rho}_1 = \rho + O(\frac{1}{\sqrt{N}}) \qquad (22)$$

Similarly,

$$\bar{\rho}_2 = \rho + O(\frac{1}{\sqrt{N}}) \qquad (23)$$

hence

$$\bar{\rho}_2 - \bar{\rho}_1 = k(\frac{1}{\sqrt{N}}), \qquad k > 0$$

From (18)

$$(h_1 - h_2)\bar{\rho}_1 \leq N(\bar{\rho}_1 - \bar{\rho}_2) + \bar{\rho}_1 [\sum_{i=1}^{N} \ell n ||H_1(e^{j\lambda_i}, \bar{\underline{\phi}}_1)||^2$$

$$- \sum_{i=1}^{N} \ell n ||H_2(e^{j\lambda_i}, \bar{\underline{\phi}}_2)||^2] + \bar{\rho}_1(n_1 - n_2)\ell nN + \xi_1\bar{\rho}_1 \qquad (24)$$

$$(h_1 - h_2)\bar{\rho}_1 \leq -\sqrt{N} \, k + [\sum_{i=1}^{N} \ell n ||H_1(e^{j\lambda_i}, \bar{\underline{\phi}}_1)||^2$$

$$- \sum_{i=1}^{N} \ell n ||H_2(e^{j\lambda_i}, \bar{\underline{\phi}}_2)||^2] + (n_1 - n_2)\bar{\rho}_1\ell nN + \xi_1\bar{\rho}_1 \qquad (25)$$

Let $\eta_2 = [\sum_{i=1}^{N} \ell n ||H_1(e^{j\lambda_i}, \bar{\underline{\phi}}_1)||^2 - \sum_{i=1}^{N} \ell n ||H_2(e^{j\lambda_i}, \bar{\underline{\phi}}_2)||^2]$

$$+ \bar{\rho}_1(n_1 - n_2)\ell nN + \xi_1\bar{\rho}_1$$

be a random variable with variance σ_2^2.

Hence

$$\text{Prob}[h_1 > h_2 | C_1]$$

$$\leq \text{Prob}[+\sqrt{N}k \leq n_2] = \text{Prob}[n_2 \geq \sqrt{N}k]$$

$$\leq O(\sigma_2^2/Nk) = O(k_2/N), \qquad k_2 > 0 \qquad (27)$$

by using the Chebychev inequality.

References

1. H. Akaike, "A new look at statistical model identification," _IEEE Trans. Automat. Contr._, Vol. AC-19, pp. 716-723, 1974.

2. R. L. Kashyap, "A Bayesian comparison of different classes of dynamic models using empirical data," _IEEE Trans. Automat. Contr._, Vol. AC-22, pp. 715-727, 1977.

3. R. L. Kashyap and A. R. Rao, _Dynamic Stochastic Models from Empirical Data_, New York: Academic Press, 1976.

4. P. Whittle, "On stationary processes in the plane," _Biometrika_, Vol. 41, pp. 434-449, 1954.

5. G. Schwarz, "Estimating the dimensions of a model," _Ann. Statistics_, Vol. 6, pp. 461-464, 1978.

6. D. R. Brillinger, _Time Series Data Analysis and Theory_, Holt, Rinehart and Winston, New York, 1974.

7. D. R. Brillinger, "Fourier analysis of stationary processes," _Proc. IEEE_, Vol. 62, pp. 1628-1642, 1974.

8. N. K. Gupta and R. K. Mehra, "Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations," _IEEE Trans. Automat. Contr._, Vol. AC-19, pp. 774-783, 1976.

9. W. E. Larimore, "Statistical inference on stationary random fields," _Proc. IEEE_, Vol. 65, pp. 961-970, 1977.

10. L. C. Hsu, "On the asymptotic evaluation of a class of multiple integrals involving a parameter," _Amer. Jl. of Math._, Vol. 73, pp. 625-634, 1951.

11. R. L. Kashyap, Lecture notes (unpublished).

12. V. S. Huzurbazar, _Sufficient Statistics_, Marcel Dekker, Inc., New York and Basel, 1976.

13. P. Whittle, "Hypothesis testing in time series analysis," thesis, Uppsala Univ., Almquist and Wiksell, Uppsala, Sweden, 1950.

14. B. H. McCormick and S. N. Jayaramamurthy, "Time series model for texture synthesis," _Int. Jl. of Computer and Inf. Sciences_, Vol. 3, pp. 329-343, 1974.

15. E. J. Delp et al., "Image modeling with a seasonal auto-regressive time series with application to data compression," Proc. Conf. PRIP-78.

16. J. T. Tou et al., "Pictorial texture analysis and synthesis," Proc. Third Int. Jt. Conf. Patt. Recn., 1976.

17. J. T. Tou and Y. S. Chang, "An approach to texture pattern analysis and recognition," Proc. IEEE Conf. on Decision and Control, 1976.

18. R. L. Kashyap, "Two-dimensional autoregressive models for images; parameter estimation and choice of neighbors," Proc. IEEE Comput. Soc. Workshop on Patt. Recn. and Art. Intelligence, 1978.

19. W. K. Pratt et al., "Visual discrimination of stochastic texture fields," IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-8, 1978, pp. 796-804.

20. W. K. Pratt and O. D. Faugeras, "Development and evaluation of stochastic-based visual texture features," Proc. Fourth Int. Jt. Conf. Patt. Recn., 1978.

21. A. Gagalowicz, "Analysis of textures using a stochastic model, ibid.

22. A. K. Jain, "Partial differential equations and finite difference methods in image processing, part 1, image representation," Jl. of Opt. Theory and Applns., Vol. 23, Sept. 1977.

23. K. Deguchi and I. Morishita, "Texture characterization and texture-based image partitioning using two-dimensional linear estimation techniques," IEEE Trans. Computers, Vol. C-27, August 1978.

24. M. Hassner and J. Slansky, "Markov random files as models of digitized image texture," Proc. Conf. PRIP-78.

25. M. Hassner and J. Slansky, "Markov random field models of digitized image texture," Proc. Fourth Int. Conf. Patt. Recn., 1978.

26. R. L. Kashyap, "Inconsistency of the AIC rule for time series models" (submitted).

27. G. E. P. Box, "Some theorems on quadratic forms applied to the study of analysis of variance problems, Ann. Math. Stat. Vol. 25, 1954, 290-302.

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* STATISTICAL INFERENCE THEORY APPLIED TO IMAGE MODELING.		5. TYPE OF REPORT & PERIOD COVERED Technical Report.
	CSC-TR-745	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Ramalingam Chellappa & Narendra Ahuja		8. CONTRACT OR GRANT NUMBER(s) AFOSR-77-3271
9. PERFORMING ORGANIZATION NAME AND ADDRESS Computer Science Center University of Maryland College Park, MD 20742		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Math. & Info. Sciences, AFOSR/NM Bolling AFB Washington, DC 20332		12. REPORT DATE March 1979
		13. NUMBER OF PAGES 49 p.
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Image processing
Pattern recognition
Image modeling

Statistical inference
Time series

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

This paper considers the application of system identification techniques using spectral representation for fitting models to textures and images and consists of two parts. In part I, we develop consistent decision rules for choosing the neighborhood in a one-dimensional autoregressive (AR) model. In part II, the theory is extended to the case of stationary two-dimensional random fields.

DD ₁ FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
403078

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*